

Acquisition de connaissances morphologiques dans le domaine médical

Pierre Zweigenbaum

CRIM, INaLCO

STIM/DSI, Assistance Publique – Hôpitaux de Paris, France

ERM 202, INSERM





2.



3. ↑



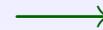
2.



3. ↑



2.



3. ↑





2.



3. ↑



4. Aspects multilingues

Plan

- ☞ Morphologie (médicale)
 - Outils et applications
 - Acquisition de connaissances morphologiques
 - Aspects multilingues

Flexion variation des mots selon les règles de la grammaire

{pied, pieds}, {généralisé, généralisée}

Dérivation construction de mots : base + affixe

dentaire; {abdomen, abdominal}

désintoxication/N [[[dé [in [toxique A] (er) V]] A] tion N]

Composition construction de mots;

composition « néoclassique » (« savante »)

adénoacanthome achondrogenèse abêtalipoprotéïnémie

Importance particulière dans un domaine technique

Dérivés et composés : exemples

L'examen doit apprécier l'état général (perte de poids) et cardio-vasculaire (hypertension artérielle), le degré de vigilance (recherche d'une confusion mentale), et les déficits sensoriels (visuel ou auditif) et moteurs pouvant interférer avec la passation des tests neuropsychologiques. L'examen neurologique reste longtemps normal dans la maladie d'Alzheimer. L'existence de signes neurologiques (signe de Babinski, syndrome pseudo-bulbaire, réflexes archaïques, signes extra-pyramidaux, troubles de la verticalité du regard, troubles sphinctériens, troubles de la posture et de la marche) doit faire évoquer un autre diagnostic que celui de maladie d'Alzheimer à l'origine des troubles cognitifs.^a

^a ANAES (2000), Recommandations pratiques pour le diagnostic de la maladie d'Alzheimer

Dérivés et composés : exemples

L'examen doit apprécier l'état général (perte de poids) et cardio-vasculaire (hypertension artérielle), le degré de vigilance (recherche d'une confusion mentale), et les déficits sensoriels (visuel ou auditif) et moteurs pouvant interférer avec la passation des tests neuropsychologiques. L'examen neurologique reste longtemps normal dans la maladie d'Alzheimer. L'existence de signes neurologiques (signe de Babinski, syndrome pseudo-bulbaire, réflexes archaïques, signes extra-pyramidaux, troubles de la verticalité du regard, troubles sphinctériens, troubles de la posture et de la marche) doit faire évoquer un autre diagnostic que celui de maladie d'Alzheimer à l'origine des troubles cognitifs.^a

^a ANAES (2000), Recommandations pratiques pour le diagnostic de la maladie d'Alzheimer

Morphologie et paliers linguistiques

- ✓ Intrinsèquement liée au lexique
(propriétés des mots)
- ✓ liée à la syntaxe
(catégorie syntaxique, nombre, genre, personne...)
sphinctériens : Adj, pluriel, masculin
- ✓ liée au sens (construction du sens)
appendicite = **appendic** + **ite**
= inflammation de l'appendice

Plan

Morphologie (médicale)

☞ Outils et applications

Acquisition de connaissances morphologiques

Aspects multilingues

Outils de traitement de la morphologie

✓ lemmatisation (FLEMM, *N*TEX)

entrée		sortie
intersectorielle/ADJ	ADJ :A-fs-	intersectoriel
pratiques/NOM	NOM :Nc-p-	pratique

Outils de traitement de la morphologie

- ✓ passage d'un dérivé affixal à sa base (DériF)
désintoxication → [[[dé [in [toxique A] (er) V]] A] tion N]
- ✓ passage d'un composé néoclassique à ses éléments constitutifs

gastralgie [[gastr N*] [algie N*] N],
(Partie de | Type particulier de) **douleur** en
relation avec le(s) **estomac**

gastrodynie [[gastr N*] [odynie N*] N],
(Partie de | Type particulier de) **douleur** en
relation avec le(s) **estomac**

Applications : recherche d'information

Recherche en texte intégral, requêtes libres

traitement	requête	extrait de texte
brut	sténoses des artères	... une artère sténosée ...
lemmatisation	sténose de <u>artère</u>	... un <u>artère</u> sténosé ...
analyse dérivés	<u>sténose</u> de <u>artère</u>	... un <u>artère</u> <u>sténose</u> ...

Applications : Indexation contrôlée

Indexation contrôlée : les termes d'index sont prédéterminés

- ▣ Indexation bibliographique (MeSH)

MeSH : INTERACTIONS ALIMENT-MEDICAMENT

Usager (lors d'une recherche) :

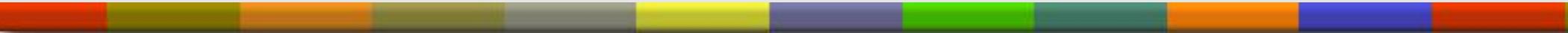
interaction entre médicaments et aliments

(alimentation | aliment | alimentaire | alimentaires) AND

(interaction | interactions) AND (médicaments | médicament)

- ▣ Enregistrement de l'activité hospitalière (CIM-10)

infarctus aigu du myocarde



Plan

Morphologie (médicale)

Outils et applications

☞ Acquisition de connaissances morphologiques

Aspects multilingues

Constituer des bases de données morphologiques

- Expertise humaine (linguiste)
- Assistance automatique (outils de TAL)
 - relation morphologique = similarité de forme + lien sémantique
 - ✍ Découverte en corpus : cooccurrence et liens thématiques
 - ✍ Découverte dans des terminologies structurées : liens sémantiques

Liens en corpus : proximité thématique

- ❶ Continuité thématique : le thème du discours ne change pas à chaque phrase ;
- ❷ Liens thématiques lexicaux (redondance lexicale) : les mots employés pour parler d'un thème donné ont souvent des liens thématiques ;
- ❸ Familles morphologiques : ces liens thématiques peuvent être instanciés par des mots d'une même famille morphologique.

➡ Conséquence :

 Parmi les mots employés à l'intérieur d'un segment de texte thématiquement homogène, on trouve souvent des mots d'une même famille morphologique.

Liens en corpus : proximité thématique

- ① Continuité thématique : le thème du discours ne change pas à chaque phrase ;
- ② Liens thématiques lexicaux (redondance lexicale) : les mots employés pour parler d'un thème donné ont souvent des liens thématiques ;
- ③ Familles morphologiques : ces liens thématiques peuvent être instanciés par des mots d'une même famille morphologique.

➡ Conséquence :

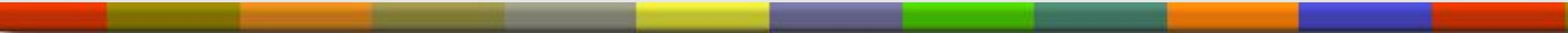
 Parmi les mots employés à l'intérieur d'un segment de texte thématiquement homogène, on trouve souvent des mots d'une même famille morphologique.

Alignement de mots dans des termes synonymes

Code	Classe	Terme anglais
P1-B6160	01	abdominal paracentesis
P1-B6160	02	abdominal tap
P1-B6160	02	abdominocentesis
P1-B6160	02	celiocentesis
P1-B6160	02	drainage of abdomen
P1-B6160	02	peritoneocentesis
P1-B6800	01	abdomen : surgical repairs, closures and reconstructions
P1-B6800	02	repair of abdominal wall, nos
P1-B6800	02	abdominoplasty
D4-D0150	01	situs inversus abdominalis
D4-D0150	02	abdominal transposition
D4-D0150	02	transposition of abdominal organs
D4-D0150	02	transposition of abdominal viscera

Alignement de mots dans des termes synonymes

Code	Classe	Terme anglais
P1-B6160	01	<u>abdominal</u> paracentesis
P1-B6160	02	<u>abdominal</u> tap
P1-B6160	02	<u>abdominocentesis</u>
P1-B6160	02	celiocentesis
P1-B6160	02	drainage of <u>abdomen</u>
P1-B6160	02	peritoneocentesis
P1-B6800	01	<u>abdomen</u> : surgical repairs, closures and reconstructions
P1-B6800	02	repair of <u>abdominal</u> wall, nos
P1-B6800	02	<u>abdominoplasty</u>
D4-D0150	01	situs inversus <u>abdominalis</u>
D4-D0150	02	<u>abdominal</u> transposition
D4-D0150	02	transposition of <u>abdominal</u> organs
D4-D0150	02	transposition of <u>abdominal</u> viscera



Plan

Morphologie (médicale)

Outils et applications

Acquisition de connaissances morphologiques



Aspects multilingues

Flexion et dérivation

- ✓ Importance variable de la flexion et de la dérivation selon les langues :

russe : flexion ++

français : flexion +

anglais : flexion –

vietnamien, chinois : flexion = 0, dérivation = 0

- ⇒ intérêt variable des outils de lemmatisation et d'analyse des dérivés

- ✓ Problématique variable de la composition
« morphologique » selon les langues :

allemand : dans la langue « générale »

Hauptbahnhof

français : dans le discours spécialisé

neuropsychologiques

Composition

- ✓ chinois : la frontière entre composés « morphologiques » et composés « terminologiques » est moins claire (pas d'espaces entre les mots)



- ➡ la problématique de la décomposition des mots composés se mêle à celle de la segmentation d'un énoncé en mots

Composés néoclassiques : parallélisme

Parallélisme des composés néoclassiques médicaux dans les langues européennes

- hépatomégalie [BAQ] HEPATOMEGALIA ; [DAN] Hepatomegali ; [DUT] Hepatomegalie ; [ENG] Hepatomegaly ; [FIN] hepatomegalia ; [FRE] HEPATOMEGALIE ; [GER] Hepatomegalie ; [HUN] hepatomegalia ; [NOR] FORSTORRET ; [POR] HEPATOMEGALIA ; [SPA] HEPATOMEGALIA

Le fonds gréco-latin est commun

- ▣ détecter les « cognats »

Composés néoclassiques : parallélisme

Disponibilité de terminologies parallèles

malignant tumor tumeur maligne

neoplasma malignes neoplasia maligna neoplasia maligna

- ▣▣▣▣ ➔ alignement de mots dans des termes alignés
(cf l'alignement de mots dans des corpus parallèles)

Généralisabilité des méthodes d'acquisition

Acquisition de connaissances flexionnelles et dérivationnelles

- ➡ L'applicabilité des méthodes présentées précédemment dépend des propriétés de la langue étudiée

Expériences en

 français

 anglais

 russe